

Discriminant Analysis of Time Series in the Presence of Within-Group Spectral Variability

Robert T. Krafty

Department of Statistics, Temple University, Philadelphia, PA, USA 19122

and

Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA 15213

rkrafty@pitt.edu

Abstract. Many studies record replicated time series epochs from different groups with the goal of using frequency domain properties to discriminate between the groups. In many applications, there exists variation in cyclical patterns from time series in the same group. Although a number of frequency domain methods for the discriminant analysis of time series have been explored, there is a dearth of models and methods that account for within-group spectral variability. This article proposes a model for groups of time series in which transfer functions are modeled as stochastic variables that can account for both between-group and within-group differences in spectra that are identified from individual replicates. An ensuing discriminant analysis of stochastic cepstra under this model is developed to obtain parsimonious measures of relative power that optimally separate groups in the presence of within-group spectral variability. The approach possess favorable properties in classifying new observations and can be consistently estimated through a simple discriminant analysis of a finite number of estimated cepstral coefficients. Benefits in accounting

for within-group spectral variability are empirically illustrated in a simulation study and through an analysis of gait variability.

Keywords. Cepstral Analysis. Fisher's Discriminant Analysis. Replicated Time Series. Spectral Analysis.

1 Introduction

Discriminant analysis of time series is important in a variety of fields including physics, geology, acoustics, economics, and medicine. In applications where scientifically meaningful information is contained within power spectra, spectral domain approaches are desired. A number of spectral based methods for the discriminant and classification analysis of time series have been developed. These methods, a review of which can be found in Chapter 7.7 of Shumway and Stoffer (2011), include Shumway and Unger (1974), Dargahi-Noubary and Laycock (1981), Shumway (1982), Alagón (1989), Zhang and Taniguchi (1994), and Kakizawa et al. (1998).

The aforementioned methods are based on models that assume the existence of group-common power spectra that can be consistently estimated from any single time series replicate from within a group. In many applications, this assumption does not hold as there exist obvious differences in second-order spectra that can be identified by individual time series within the same group. As an example, consider stride interval series, or the time taken to complete consecutive gait cycles, that were collected as part of a study to better understand connections between walking patterns and neurological disease (Hausdorff et al., 2000). Figure 1 displays three examples of stride interval series from study participants under each of three neurological conditions and Figure 2 displays estimated log-spectra for each stride interval series in the study. Aside from neurological conditions, walking is influenced by numerous person specific physiological characteristics. Consequently, there exist differences in cyclical patterns of stride interval series from different subjects with a common neurological condition.

The presence of extra spectral variability in replicated time series and the inability of traditional time series models and methods to account for it was first discussed in the literature by Diggle and Al Wasel (1997). They introduced a parametric log-spectral mixed-effects model for replicated time series, which was later nonparametrically generalized by Saavedra et al. (2000, 2008), Iannaccone

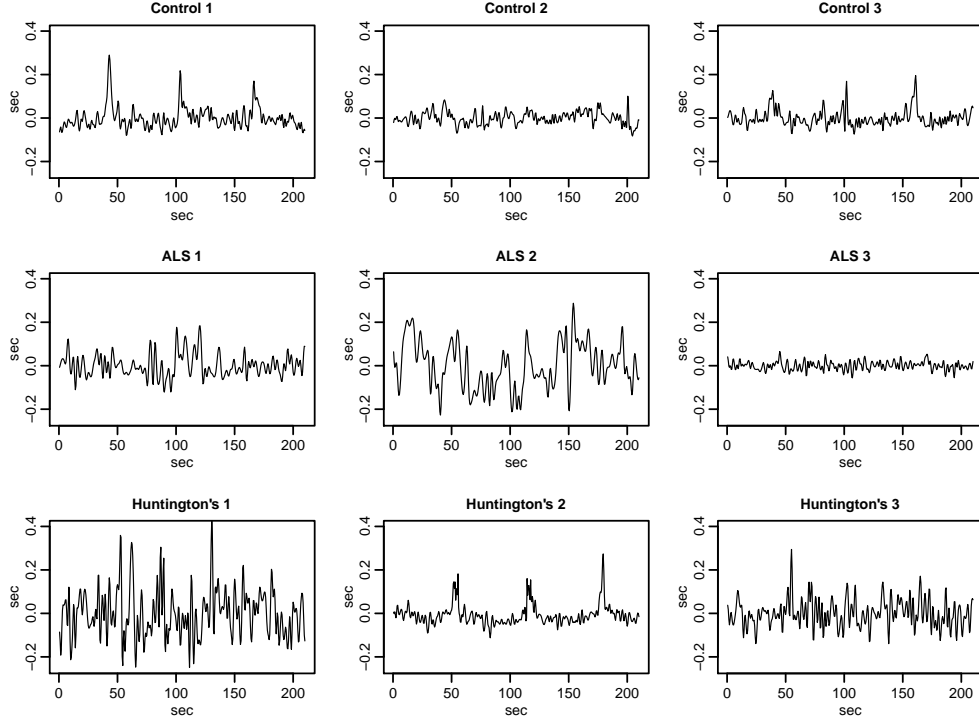


FIGURE 1: Detrended stride interval series from 9 participants in the gait analysis study: 3 healthy controls, 3 participants with amyotrophic lateral sclerosis (ALS), and 3 participants with Huntington's disease.

and Coles (2001), Freyermuth et al. (2010) and Krafty et al. (2011). Although these models allow one to conduct inference on a group average spectrum, the question of how within-group spectral variability affects discriminant analysis has yet to be addressed. To address this question, this article introduces a stochastic transfer function model for groups of time series in the presence of within-group spectral variability that makes explicit the replicate-specific spectra that are estimable from individual series. The model accounts for the higher-order long-range dependence that is responsible for within-group spectra variability.

The Fisher's discriminant analysis of stochastic cepstra, or inverse Fourier transforms of log-spectra, is explored as a means of discriminating and classifying time series under the stochastic

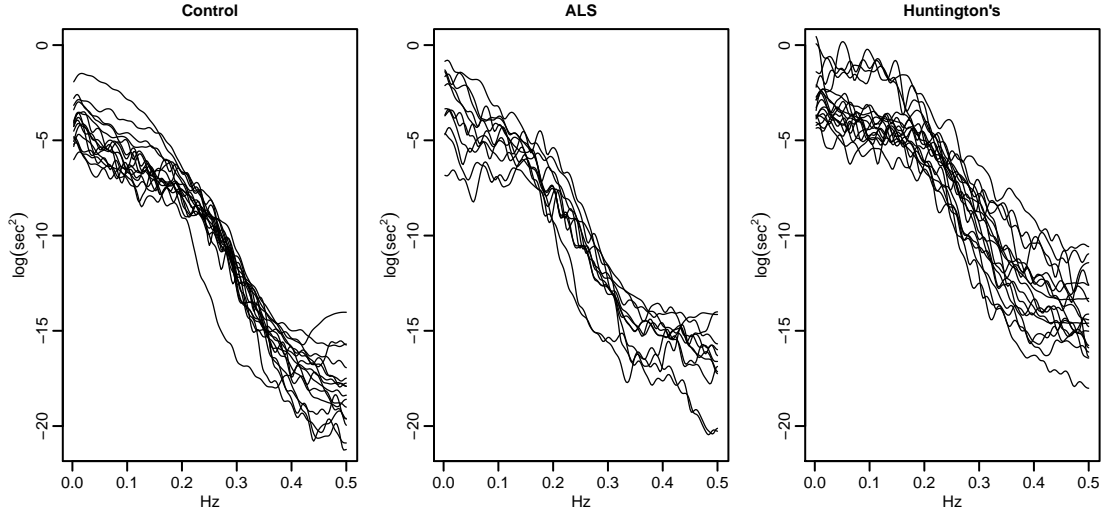


FIGURE 2: Estimated replicate specific log-spectra for the detrended stride interval series from each participant in the gait analysis study.

transfer function model. The procedure possess optimal properties that mirror those of Fisher-type discriminants in other high-dimensional settings. When extra spectral variability exists, the cepstral Fisher's discriminant analysis more accurately classifies a new time series of unknown group membership compared to existing spectral methods that ignore within-group spectral variability. Other Fisher's type approaches can be developed under the stochastic transfer function model, such as through integral functions of log-spectra. We use the cepstral formulation as it provides both a more encompassing theoretical framework and an intuitive approach to estimation that simultaneously overcomes the inconsistency of periodograms and the high-dimensionality of the data.

As discussed in Chapter 11 of Johnson and Wichern (2007), although intertwined, there is a distinction between discriminant analysis, which seeks parsimonious measures that illuminate group separation, and classification analysis, or the prediction of group membership of new observations. The aforementioned existing methods for spectral discrimination address the problem of classifica-

tion, albeit with a higher error than the proposed procedure when within-group variability exists, but are black-box in nature and do little in helping to understand the way in which groups are separated. The proposed procedure address the discrimination problem by producing discriminants and weight functions that provide low-dimensional interpretable measures of relative power that illustrate scientific mechanisms that separate groups.

The rest of this article is organized as follows. The stochastic transfer function model for groups of replicated time series with within-group spectral variability is presented in Section 2. In Section 3, we discuss the cepstral Fisher’s procedure for discriminant and classification analysis. Section 4 introduces an intuitive estimation procedure via a finite number of estimated cepstral coefficients. Simulation studies are explored in Section 5 to investigate empirical properties of the proposed procedure and to compare these properties to those of existing methods. The method is used in Section 6 to analyze data from the motivating study of gait variability. Concluding remarks are given in Section 7. Proofs are relegated to the appendix.

2 Model

2.1 Stochastic Transfer Function Model

Consider a population of real-valued stationary time series composed of $j = 1, \dots, J$ groups, Π_j represents the j th group, and π_j is the proportion of time series from the population belonging to Π_j . To present a model that accounts for both between and within-group spectral variability, we consider a model with stochastic replicate-specific transfer functions. Replicate-specific transfer functions A_{jk} are defined as independent second-order random variables identically distributed for each replicate k within group j and whose realizations are complex-valued random functions that are Hermitian, have period 1, and both real and imaginary parts are uniformly continuous. The

group Π_j is modeled as the collection of time series $\{X_{jkt}; t \in \mathbb{Z}\}$ where

$$X_{jkt} = \int_0^1 A_{jk}(\lambda) e^{2\pi i \lambda t} dZ_{jk}(\lambda), \quad (2.1)$$

Z_{jk} are independent and identically distributed orthogonal processes that are independent of the transfer functions and $E |dZ_{jk}(\lambda)|^2 = 1$. Mixing conditions are assumed on Z_{jk} such that $z_{jkt} = \int_0^1 e^{2\pi i \lambda t} dZ_{jk}(\lambda)$ is, in some sense, short-range dependent. Many different mixing conditions can be used. In this article we use the conditions in Assumption 2.6.1 of Brillinger (2001) such that z_{jkt} is strictly stationary, cumulants of all orders exists and are absolutely summable.

Each time series in Π_j is an independent and identically distributed stationary process with Cramér representation $\int_0^1 e^{2\pi i \lambda t} dW(\lambda)$, $W(\lambda) = \int_0^\lambda A_{jk}(\omega) dZ_{jk}(\omega)$, and spectrum $E|A_{jk}(\lambda)|^2$. When $\text{Var} \left\{ |A_{jk}(\lambda)|^2 \right\} = 0$ for all λ , X_{jkt} is short-range dependent. However, when the variance of replicate-specific spectra is non-trivial, it is not. Conditional on the replicate-specific random transfer function A_{jk} , X_{jkt} is stationary, short-range dependent, and has replicate-specific power spectrum $|A_{jk}(\lambda)|^2$. The stochastic transfer function model is a reparameterization of the Cramér representation that makes explicit the unit-specific spectra that are estimable from individual replicates. Before comparing the stochastic transfer function model to other models, consider the following example.

Example: Conditional MA(1). Let Π_j be the collection of conditional invertible Gaussian MA(1) processes with nonnegative autocorrelation such that $X_{jkt} = \epsilon_{jkt} + \theta_{jk} \epsilon_{jkt-1}$ where θ_{jk} are uniformly distributed over $[0, 1]$ and are independent of $\epsilon_{jkt} \stackrel{iid}{\sim} N(0, \sigma^2)$. These time series have a stochastic transfer function representation of the form (2.1) where $A_{jk}(\lambda) = \sigma (1 + \theta_{jk} e^{-2\pi i \lambda})$ and Z_{jk} is complex Brownian motion with $\text{Cov} \{Z_{jk}(\lambda), Z_{jk}(\omega)\} = \min(\lambda, \omega)$ and $Z_{jk}(-\lambda) = \overline{Z_{jk}(\lambda)}$. Conditional on θ_{jk} , X_{jkt} is a Gaussian MA(1) process with replicate-specific spectrum $|A_{jk}(\lambda)|^2 = \sigma^2 \left\{ 1 + \theta_{jk}^2 + 2\theta_{jk} \cos(2\pi\lambda) \right\}$. As a Gaussian MA(1) process, the conditional autocovariance and all conditional higher-order moments vanish for lags greater than 1. Consequently, standard results

hold so that the periodogram from a replicate at Fourier frequencies are approximately independent and can be smoothed to obtain a consistent estimate of the replicate-specific spectrum. Marginally, X_{jkt} is stationary, $\text{Cov}(X_{jkt+h}, X_{jkt}) = 0$ when $|h| > 1$, and it has a marginal power spectrum of $E|A_{jk}(\lambda)|^2 = \sigma^2 \{4/3 + \cos(2\pi\lambda)\}$. However, higher order cumulants do not decay. For instance, $\text{Cov}(X_{jkt+h}^2, X_{jkt}^2) = 4\sigma^4/45$ for all $|h| > 1$. Consequently, periodograms are correlated and cannot be smoothed to obtain a consistent estimate of the marginal spectrum.

The existing models for spectral based discriminant analysis previously cited are based on the assumption that each time series is short-range dependent so that periodograms at different Fourier frequencies are approximately independent, and that periodograms from each realization can be smoothed to obtain a consistent estimate of a group-common spectrum. Such assumptions are required in the spectrum analysis of a single time series, as they are essential in obtaining a consistent estimate. However, they are not appropriate for the analysis of replicated time series when there exists variability in the second-order spectral structures from different realizations. The stochastic transfer function model generalizes the models considered by these existing methods to account for the marginal higher-order long-range dependence that is responsible for within-group spectral variability. It should be noted that stochastic transfer functions were previously used in the semiparametric mixed-effects regression model of Krafty et al. (2011), where transfer functions are decomposed into deterministic and stochastic components to allow for the regression analysis of time series when multiple correlated replicates are observed from different subjects. Although a parameterization of this model using dummy variables can be formulated to define a model of groups of independent time series with extra spectral variability, it is not identifiable.

2.2 Stochastic Cepstra and Log-Spectra

Our goal is to find interpretable measures that best separate groups and focus on those that are measures of relative power, or are linear in log-spectra. To find such measures, we will utilize cepstral coefficients, or inverse Fourier transforms of log-spectra. They will provide both a flexible and rigorous theoretical framework and guide estimation. Replicate-specific log-spectra are defined as $\gamma_{jk}(\lambda) = \log |A_{jk}(\lambda)|^2$, the j th group mean log-spectrum is defined as $\alpha_j(\lambda) = E\{\gamma_{jk}(\lambda)\}$, and we let $\beta_{jk}(\lambda) = \gamma_{jk}(\lambda) - \alpha_j(\lambda)$ be the replicate-specific deviation of the log-spectrum of the k th replicate from the j th group. The original formulation of the cepstrum by Bogert et al. (1963) considered coefficients in terms of complex trigonometric polynomials. Since we are considering real-valued time series, log-spectra are even functions and, in a manner similar to Bloomfield (1973), we define the cepstrum through a cosine series. Define the replicate-specific cepstrum as $c_{jk} = (c_{jk0}, c_{jk1}, \dots) \in \mathbb{R}^{\mathbb{N}}$, where $\mathbb{R}^{\mathbb{N}}$ is the set of real valued sequences indexed by the natural numbers $\mathbb{N} = \{0, 1, \dots\}$, such that

$$c_{jk0} = \int_0^1 \gamma_{jk}(\lambda) d\lambda, \quad c_{jk\ell} = \int_0^1 \gamma_{jk}(\lambda) \sqrt{2} \cos(2\pi\lambda\ell) d\lambda, \quad \ell = 1, 2, \dots$$

Group-average and replicate-specific deviation cepstra $a_j, b_{jk} \in \mathbb{R}^{\mathbb{N}}$ are similarly defined as the cosine series of α_j and β_{jk} , respectively.

Example: Conditional MA(1), continued. For the collection of conditional MA(1) processes considered in Section 2.1, replicate specific log-spectra are

$$\gamma_{jk}(\lambda) = \log [\sigma^2 \{1 + \theta_{jk}^2 + 2\theta_{jk} \cos(2\pi\lambda)\}] = \log(\sigma^2) + 2 \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1} \theta_{jk}^{\ell}}{\ell} \cos(2\pi\ell\lambda),$$

so that $c_{jk0} = \log(\sigma^2)$, $c_{jk\ell} = (-1)^{\ell+1} \sqrt{2} \theta_{jk}^{\ell} / \ell$, $\ell \geq 1$. Recalling that θ_{jk} is uniformly distributed over the unit interval, the group-average cepstra can be found to be $a_{jk0} = \log(\sigma^2)$, $a_{jk\ell} = (-1)^{\ell+1} \sqrt{2} / \{\ell(\ell+1)\}$, $\ell \geq 1$. The zero-order cepstral coefficient reflects the conditional innovation variance. In our example, replicates have a common conditional innovation variance, so

that $\text{Var}(c_{jk0}) = 0$. Within-group variability in smooth replicate-specific spectra is reflected in the variability of positive-order cepstral coefficients, with

$$\text{Cov}(c_{jk\ell}, c_{jkm}) = \frac{2(-1)^{\ell+m+2}}{(\ell+m+1)(\ell+1)(m+1)}, \quad \ell, m \geq 1.$$

3 Discriminant and Classification Analysis

3.1 Cepstral Discriminant Analysis

The cepstral Fisher's discriminant analysis seeks successive uncorrelated one-dimensional linear functions of replicate-specific cepstra that provide maximum separation between group-means relative to within-group variability. Before defining the procedure, we first define some notation. Let $\bar{a}_\ell = \sum_{j=1}^J \pi_j a_{j\ell}$ be the overall mean cepstrum and define the between-group kernel as $\Lambda(\ell, m) = \sum_{j=1}^J \pi_j (a_{j\ell} - \bar{a}_\ell)(a_{jm} - \bar{a}_m)$. Separation between group-means of linear functions of cepstra $\sum_{\ell=0}^{\infty} y_{0\ell} c_{jk\ell}$ with weights $y_0 \in \mathbb{R}^{\mathbb{N}}$ is defined as the weighed sum of the squared distances of each group-mean to the overall mean, or $\|y_0\|_{\Lambda}^2 = \sum_{\ell, m=0}^{\infty} y_{0\ell} \Lambda(\ell, m) y_{0m}$. Defining the within-group kernel $\Gamma(\ell, m) = E(b_{jk\ell} b_{jkm})$, the covariance between two linear combinations of cepstra with weights $y_0, y_1 \in \mathbb{R}^{\mathbb{N}}$ can be written as $\langle y_0, y_1 \rangle_{\Gamma} = \sum_{\ell, m=0}^{\infty} y_{0\ell} \Gamma(\ell, m) y_{1m}$ and the variance of a linear combination with weights y_0 is given by $\|y_0\|_{\Gamma}^2 = \langle y_0, y_0 \rangle_{\Gamma}$.

Discriminants are defined sequentially. First discriminants d_{jk1} are linear functions of cepstra defined by weights $y_1 \in \mathbb{R}^{\mathbb{N}}$ such that group means are maximally separated relative to within-group variability such that

$$d_{jk1} = \sum_{\ell=0}^{\infty} y_{1\ell} c_{jk\ell}, \quad y_1 = \underset{\|y\|_{\Gamma}=1}{\operatorname{argmax}} \|y\|_{\Lambda}.$$

Higher order discriminants maximize group-mean separation among linear combinations orthogonal to lower-order discriminants such that q th-order discriminants d_{jkq} with weights $y_q \in \mathbb{R}^{\mathbb{N}}$ are defined

as

$$d_{jkq} = \sum_{\ell=0}^{\infty} y_{q\ell} c_{jk\ell}, \quad y_q = \underset{\|y\|_{\Gamma}=1, \langle y, y_m \rangle_{\Gamma}=0, m < q}{\operatorname{argmax}} \|y\|_{\Lambda}.$$

The number of non-trivial discriminants Q is less than or equal to the ranks of Γ and Λ , which is less than or equal to $J - 1$. The discriminants d_{jkq} provide low-dimensional measures that best separate the J groups. These parsimonious measures can be used to visualize the high-dimensional data, such as Figure 5 in the analysis of gait variability in Section 6, and provide an intuitive a powerful classification procedure, which is discussed in Section 3.3. The weights y_q provide information as to how the discriminants can be interpreted as measures of relative power.

The cepstral Fisher's discriminant analysis is formulated above when the covariance functions of b_{jk} are the same for each group. If there is heterogeneity, all statements made in this article aside from those concerning optimal classification rates still apply using the pooled within-group covariance $\sum_{j=1}^J \pi_j E(b_{jk\ell} b_{jkm})$ in lieu of $\Gamma(\ell, m)$. A more detailed discussion of this issue is presented in Section 7. A consequence of Proposition 2.1 from Shin (2008) is that cepstral discriminants are well defined when cepstral group means are not dissimilar to replicate-specific deviations in the sense that a_j is contained in the reproducing kernel Hilbert space with reproducing kernel Γ for $j = 1, \dots, J$. We assume that this regularity condition is satisfied so that discriminants always exist.

3.2 Log-Spectral Weight Functions and Discriminants

Measures of relative power could have alternatively been defined using integral functions of log-spectra. However, due to the unbounded nature of the inverse of a non-singular continuous covariance operator, optimal discriminants will not necessarily exist unless additional strong and nonassessable assumptions are made. This issue is discussed by Shin (2008) and by Delaigle and Hall (2012). When $\sum_{\ell=0}^{\infty} |y_{q\ell}| < \infty$, $\xi_q(\lambda) = y_{q0} + \sum_{\ell=1}^{\infty} y_{q\ell} \sqrt{2} \cos(2\pi\lambda\ell)$ exists, $d_{jkq} = \int_0^1 \xi_q(\lambda) \gamma_{jk}(\lambda) d\lambda$,

and we refer to ξ_q as a log-spectral weight function. The cepstral based formulation is broader and encompass the integral log-spectral formulation in the sense that, if a discriminant exists in the integral log-spectral formulation, it is equivalent to d_{jkq} and has weight function ξ_q .

3.3 Classification Analysis

Consider a new time series of unknown group membership $\{X_{*t}; t \in \mathbb{Z}\}$ with cepstrum c_* and q th cepstral discriminant d_{*q} . The property that d_{*q} , $q = 1, \dots, Q$, are uncorrelated with unit variance suggests classifying the new time series into Π_j when the j th group-mean discriminants $\mu_{jq} = \sum_{\ell=0}^{\infty} y_{q\ell} a_{j\ell}$ are closer to $(d_{*1}, \dots, d_{*Q})^T$ than the other $J-1$ group means. Formally, if we let $\Pi(c_*) \in \{1, \dots, J\}$ index the population to which the new time series is classified, the classification rule is

$$\Pi(c_*) = \operatorname{argmin}_{j=1, \dots, J} \left\{ \sum_{q=1}^Q (d_{*q} - \mu_{jq})^2 - 2 \log(\pi_j) \right\}. \quad (3.2)$$

This classification rule is the optimal centroid classifier of Delaigle and Hall (2012) for replicate-specific log-spectra and properties concerning its classification rate are established in their Theorems 1 and 2 when Γ is non-singular and $J = 2$. If the processes generating γ_{jk} are Gaussian, then (3.2) is optimal in that it has smallest classification error among all spectrum based classification rules, and this error is bounded from zero. If γ_{jk} is not Gaussian, then the classification rule with smallest error will depend on the distribution of γ_{jk} , and this error is bounded from zero. Although a non-linear classifier will exist with smaller classification error, one must know the distribution of γ_{jk} to find this rule and, while it presents some theoretical improvements compared to the Fisher's cepstral procedure for the classification problem, it will not provide parsimonious measures to address the discrimination problem.

When within-group spectral variability is not present, asymptotically perfect classification can be achieved (Zhang and Taniguchi, 1994; Kakizawa et al., 1998). However, when within-group spec-

tral variability is present, asymptotically perfect classification is not possible and methods that can achieve asymptotically perfect classification in the absence of within-group spectral variability have a bounded non-zero error rate. As a time series from one group could possess a replicate-specific spectrum that more closely resembles replicate-specific spectra from another group by chance alone, this is rather intuitive and illustrates the increased difficulty of classification in the presence of within-group spectral variability. It should be noted that Delaigle and Hall (2012) describe a situation where asymptotically perfect classification is possible for functional data. However, the assumption that a_j is contained in the reproducing kernel Hilbert space with reproducing kernel Γ , which is known to be necessary in avoiding degenerate time series models (Parzen, 1961, 1962), makes asymptotically perfect classification not possible under the stochastic transfer function model.

4 Estimation

4.1 Cepstral Coefficients

Consider estimation from $n = \sum_{j=1}^J n_j$ independent time series epochs of length N , $\{X_{jk1}, \dots, X_{jkN}\}$, $j = 1, \dots, J$, $k = 1, \dots, n_j$. When replicate-specific log-spectra are smooth, most information is contained in lower-order cepstral coefficients. This is illustrated in the conditional MA(1) example from Sections 2.1 and 2.2, where $c_{jkl} = O_p(\ell^{-3/2})$. Consequently, discriminants and weight functions can be estimated through a classical Fisher's discriminant analysis of a finite number of estimated cepstral coefficients.

We consider the class of replicate-specific cepstral estimators of the form

$$\hat{c}_{jk0} = N^{-1} \sum_{m=0}^{N-1} \hat{\gamma}_{jkm}, \quad \hat{c}_{jkl} = N^{-1} \sum_{m=0}^{N-1} \hat{\gamma}_{jkm} \sqrt{2} \cos(2\pi \lambda_m \ell), \quad \ell = 1, \dots, \lfloor N/2 \rfloor,$$

where $\hat{\gamma}_{jkm}$ is an estimator of $\gamma_{jk}(\lambda_m)$, $\lambda_m = m/N$, $m \in \mathbb{Z}$. There is an extensive literature on

spectrum estimation. Since, for fixed ℓ , the variance of $\hat{c}_{jk\ell} - c_{jk\ell}$ decays as N increases, not only are standard consistent estimators appropriate, such as smoothed periodograms and multitaper estimates, but also inconsistent estimators, such as periodograms. Although the consistency results for discriminants and weight functions established in Section 4.4 hold for any standard estimator, good finite sample performance is contingent on log-spectral estimators having small bias, small variance and, since the within-group covariance Γ must also be estimated, errors that are approximately uncorrelated across frequency. We advocate the use of the multitaper estimators developed by Thomson (1982) that are shown in Percival and Walden (1993) to have small bias, as opposed to periodogram based methods, small variance, as opposed to inconsistent estimators, and high resolution, as opposed to methods that smooth across frequency.

Let $\{h_{rt}, t = 1, \dots, N\}$ be $r = 1, \dots, R$ non-negative orthonormal data tapers such that $\sum_{t=1}^N h_{rt}h_{st} = 1\{r = s\}$, $r, s = 1, \dots, R$, where $1\{\cdot\}$ is the indicator function. This article will consider the sine tapers

$$h_{rt} = \left(\frac{2}{N+1}\right)^{1/2} \sin\left(\pi t \frac{r}{N+1}\right).$$

The r th direct spectral estimator is defined as the tapered periodogram under the r th data taper, $I_{jkrm} = \left|N^{-1/2} \sum_{t=1}^N h_{rt} X_{jkt} e^{-2\pi i \lambda_m t}\right|^2$, and the multitaper log-spectral estimator is defined as

$$\hat{\gamma}_{jkm} = \log \left(R^{-1} \sum_{r=1}^R I_{jkrm} \right). \quad (4.3)$$

4.2 Discriminants, Weight Functions, and Classification

We estimate discriminants and weight functions through a classical Fisher's discriminant analysis of $\hat{\mathbf{c}}_{jk}^L = (\hat{c}_{jk0}, \dots, \hat{c}_{jkL-1})^T$ for some L smaller than N and n . The data driven selection of L is discussed in Section 4.3. To define this estimator, let $\hat{\pi}_j = n_j/n$, $\hat{\mathbf{a}}_j^L = n_j^{-1} \sum_{k=1}^{n_j} \hat{\mathbf{c}}_{jk}^L$, $\hat{\hat{\mathbf{a}}}^L = \sum_{j=1}^J \hat{\pi}_j \hat{\mathbf{a}}_j^L$, and $\hat{\mathbf{\Lambda}}_L = \sum_{j=1}^J \hat{\pi}_j \left(\hat{\mathbf{a}}_j^L - \hat{\hat{\mathbf{a}}}^L \right) \left(\hat{\mathbf{a}}_j^L - \hat{\hat{\mathbf{a}}}^L \right)^T$. Further, let $\hat{\mathbf{b}}_{jk}^L = \hat{\mathbf{c}}_{jk}^L - \hat{\mathbf{a}}_j^L$ and $\hat{\mathbf{\Gamma}}_L = \sum_{j=1}^J \hat{\pi}_j \hat{\mathbf{\Gamma}}_{Lj}$ where $\hat{\mathbf{\Gamma}}_{Lj} = (n_j - 1)^{-1} \sum_{k=1}^{n_j} \hat{\mathbf{b}}_{jk}^L \left(\hat{\mathbf{b}}_{jk}^L \right)^T$. An estimated q th cepstral weight function $\hat{\mathbf{y}}_q^L$ is defined as a

q th ordered eigenvector of $\hat{\mathbf{\Gamma}}_L^{-1} \hat{\mathbf{\Lambda}}_L$ and an estimated discriminant is defined as $\hat{d}_{jkq}^L = (\hat{\mathbf{y}}_q^L)^T \hat{\mathbf{c}}_{jk}^L$.

Note that the estimated log-spectral weight function $\hat{\xi}_q^L(\lambda) = \hat{y}_{q0}^L + \sum_{\ell=1}^{L-1} \hat{y}_{q\ell}^L \sqrt{2} \cos(2\pi\lambda\ell)$ exists and is interpretable even if ξ_q itself only exists in a limiting sense. The classification rule for a new time series with estimated L -dimensional cepstrum $\hat{\mathbf{c}}_*^L$ can be estimated as $\hat{\Pi}(\hat{\mathbf{c}}_*^L) = \operatorname{argmin}_j \left[\sum_{q=1}^Q \left\{ \left(\hat{\mathbf{c}}_*^L - \hat{\mathbf{a}}_j^L \right)^T \hat{\mathbf{y}}_q^L \right\}^2 - 2 \log(\hat{\pi}_j) \right]$.

It should be noted that $\hat{c}_{jk\ell} \approx c_{jk\ell}$ for all $\ell = 0, \dots, L-1$ when a reasonable spectral estimator is chosen and when L is small relative to N . Consequently, in practice, the procedure is similar to a classical Fisher's discriminant analysis on $(c_{jk0}, \dots, c_{jkL-1})^T$, and statistical properties of estimators follow from classical results for estimated Fisher's discriminant analysis (Anderson, 2003, Chapter 6).

4.3 Selecting L

A leave-out-one cross-validation procedure akin to the cross-validation commonly used for other regularized Fisher's discrimination procedures can be used for the data driven selection of L . Let $\hat{\Pi}_{[jk]}^L$ be the index of the classification rule for the k th time series from Π_j using L cepstral coefficients when weight functions are estimated using the $n-1$ time series that exclude this series. The cross-validation rule selects $L = \operatorname{argmin}_\ell \left[\sum_{j=1}^J \sum_{k=1}^{n_j} 1 \left\{ \hat{\Pi}_{[jk]}^\ell \neq j \right\} \right]$.

4.4 Consistency

Consistency of estimated discriminants, weight functions, and classification rules can be established under some assumptions as n_j , N , and L increase. First, it is assumed that $\hat{\pi}_j$ is \sqrt{n} -consistent, which holds under simple random sampling.

Assumption 1. $\hat{\pi}_j = \pi_j + O_p(n^{-1/2})$.

Appropriate estimators of π_j based on the sampling scheme can be used if this does not hold.

We will assume the following two regularity conditions so that cepstral coefficients decay at a sufficient rate and that eighth moments of X_{jkt} exist.

Assumption 2. $\sum_{\ell=1}^{\infty} \ell^2 |a_{j\ell}|^2 < \infty$ and $\Pr\left(\sum_{\ell=1}^{\infty} \ell^2 |b_{jk\ell}|^2 < \infty\right) = 1, j = 1, \dots, J$.

Assumption 3. $\sup_{\lambda \in \mathbb{R}} E\{|dZ_{jk}(\lambda)|^8\} < \infty$ and $\sup_{\ell \in \mathbb{N}} E|b_{jk\ell}|^4 < \infty, j = 1, \dots, J$.

Assumption 2 implies that mean and replicate-specific deviations of log-spectra are absolutely continuous with square integrable first derivatives and, consequently, error incurred by using only a finite number of cepstral coefficients is asymptotically negligible. The moment conditions in Assumption 3 imply that the eighth moments of X_{jkt} are bounded. Our estimators depend on $\hat{\Gamma}_L$ and $\hat{\Lambda}_L$, which are fourth-order functions of X_{jkt} , and these conditions assure their variances exist. Additionally, it is assumed that a log-spectral estimator is chosen that has asymptotically optimal bias (up to a constant), resolution, and bounded variance in the sense that the following assumption holds.

Assumption 4. For $j = 1, \dots, J$, $\sup_{m=0, \dots, \lfloor N/2 \rfloor} E\{\hat{\gamma}_{jkm} - \gamma_{jk}(\lambda_m)\} = \nu_1 + O(N^{-1/2})$, for some $\nu_1 \in \mathbb{R}$, $\sup_{m \neq p=0, \dots, \lfloor N/2 \rfloor} \text{Cov}\{\hat{\gamma}_{jkm} - \gamma_{jk}(\lambda_m), \hat{\gamma}_{jkp} - \gamma_{jk}(\lambda_p)\} = O(N^{-1})$, and $\sup_{m=0, \dots, \lfloor N/2 \rfloor} \text{Var}\{\hat{\gamma}_{jkm} - \gamma_{jk}(\lambda_m)\} = \nu_2 + O(N^{-1})$ for some $\nu_2 > 0$.

Log-spectral estimators need only be asymptotically unbiased up to a constant since the discriminant analysis is invariant to the addition of a constant. Assumption 4 holds for multitaper estimates, including tapered periodograms when $R = 1$. Whereas consistency results for multitaper estimates from a single realization require the number of tapers to grow with respect to the number of time points, we consider fixed R , since the variance reduction achieved by increasing the number of tapers is inherently achieved by projecting onto a finite cosine basis.

Lemma 1. *Assumption 4 holds for $\hat{\gamma}_{jkm}$ defined by (4.3) under Assumptions 2 and 3.*

The proof of Lemma 1 mirrors the proof of Theorem 1 from Krafty et al. (2011). It should be noted that, although this theorem is formulated for untapered periodograms, the properties underlying the Bartlett's expansion used in the proof that are formulated in Janas and von Sachs (1995) hold for tapered data.

The following theorem establishes consistency under these assumptions when the growth of L is restricted by the growth of n and N . The main obstacle in performing Fisher's discriminant analysis when the dimension is large compared to the number of observations stems from the divergence of the eigenvalues of the within-group covariance and its inverse. The smoothness of log-spectra implies that the largest eigenvalue of the truncated within-group covariance matrix $E \left\{ \mathbf{b}_{jk}^L \left(\mathbf{b}_{jk}^L \right)^T \right\}$ is bounded from above but that its smallest, which we will refer to as σ_L , approaches zero as $L \rightarrow \infty$. To assure that the inverse of $E \left\{ \mathbf{b}_{jk}^L \left(\mathbf{b}_{jk}^L \right)^T \right\}$ can be consistently estimated by its sample version, we must assume that the number of replicates grows quickly compared to the number of coefficients by assuming that $Ln^{-1/2} \rightarrow 0$ and $\sigma_L^{-2}n^{-1/2} \rightarrow 0$. Replicated-specific cepstral coefficients are not observed but estimated from replicates of length N . To ensure that the error incurred by this estimation is asymptotically negligible, we must also limit the growth of the number of coefficients by the length of the time series such that $\sigma_L^{-2}N^{-1/2} \rightarrow 0$.

Theorem 1. *Under Assumptions 1-4, for every q th weight function y_q , there exist a series of q th eigenvectors $\hat{\mathbf{y}}_q^L$ of $\hat{\mathbf{\Gamma}}_L^{-1}\hat{\mathbf{\Lambda}}_L$ such that, if $\sigma_L^{-2}n^{-1/2} \rightarrow 0$, $\sigma_L^{-2}N^{-1/2} \rightarrow 0$ and $Ln^{-1/2} \rightarrow 0$ as $n, N, L \rightarrow \infty$, $\|(\hat{\mathbf{y}}_q^L, 0, \dots) - y_q\|_{\Gamma} \xrightarrow{p} 0$.*

A direct consequence of the Γ -norm consistency of weight functions is that $\hat{d}_{jkq}^L \xrightarrow{p} d_{jkq}$ and $\hat{\Pi}(\hat{\mathbf{c}}_*^L) \xrightarrow{p} \Pi(c_*)$.

5 Simulation

We conducted a simulation study to investigate the empirical properties of classification procedures when within-group spectral variability exists. In these simulations, time series from $J = 3$ groups are simulated as conditional autoregressive processes $X_{jkt} = \phi_{jk1}X_{jkt-1} + \phi_{jk2}X_{jkt-2} + \epsilon_{jkt}$ where ϵ_{jkt} are independent Gaussian white noise with conditional variance σ_{jk}^2 . Autoregressive parameters are drawn as independent uniform random variables where $\phi_{1k1} \sim \text{Uni}(0.05, 0.7)$, $\phi_{1k2} \sim \text{Uni}(-0.12, -0.06)$, $\phi_{2k1} \sim \text{Uni}(0.01, 1.2)$, $\phi_{2k2} \sim \text{Uni}(-0.36, -0.25)$, $\phi_{3k1} \sim \text{Uni}(0.12, 1.5)$, and $\phi_{3k2} \sim \text{Uni}(-0.75, -0.56)$. Three ranges for conditional innovation variances are explored where σ_{jk}^2 are drawn as independent uniform random variables over $[0.1, 10]$, $[0.3, 3]$, and $[0.9, 1.1]$. One thousand random samples are drawn for each of the 27 combinations of the three ranges of innovation variances, three numbers of time series per group in training data $n_j = 15, 50, 100$, and three time series lengths $N = 250, 500, 1000$. Figure 3 displays simulated replicate-specific log-spectra when $n_j = 50$ and $\sigma_{jk}^2 \in [0.3, 3]$. A test data set of 50 time series per group is drawn for each random sample to evaluate out-of-sample classification rates.

The proposed cepstral Fisher’s discriminant analysis was implemented using three log-spectral estimators: the multitaper estimator with $R = 7$, a direct estimator, and a smoothed estimator. Direct estimators were taken to be first tapered log-periodograms. Smoothed estimates were obtained by taking the logarithm after smoothing I_{jk1m} across frequency via the modified Daniell smoother with span selected through generalized cross-validation (Ombao et al., 2001). Additionally, two popular methods for the discriminant analysis of time series in the absence of within-group spectral variability were implemented using Kullback-Leibler and Chernoff information measures (Kakizawa et al., 1998; Shumway and Stoffer, 2011). Replicate-specific spectra were individually estimated by smoothing periodograms with a modified Daniell smoother and span selected through generalized cross-validation. A test time series is classified into Π_j when the information measure

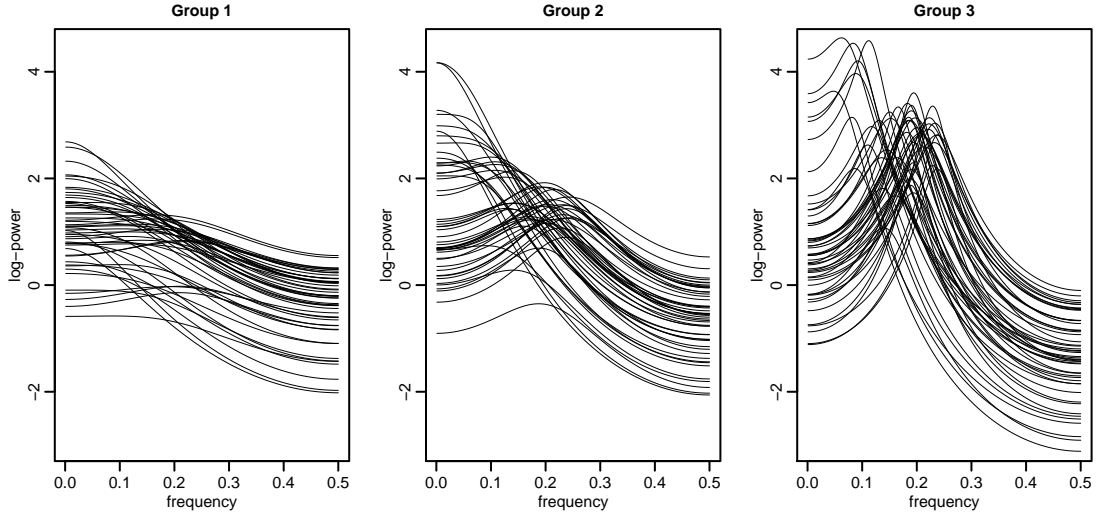


FIGURE 3: Replicate-specific log-spectra from simulated conditional AR(2) processes with $n_j = 50$ and $\sigma_{jk}^2 \in [0.3, 3]$.

distance between its smoothed periodogram and the average of the smoothed periodograms from the Π_j training data is smaller than its distance to the average of the smoothed periodograms from the training data from the other two groups. The tuning parameter for the Chernoff measure was selected using an appropriately modified version of the cross-validation procedure outline in Section 4.3.

Table 1 displays the mean and standard deviation of the classification rates. In every setting, the cepstral Fisher’s discriminant analysis, using any of the three log-spectral estimators, had higher mean classification rates than the two information criterion based methods. Although the three cepstral Fisher’s discriminant procedures display comparatively similar performance, in each setting, the multitaper based method had the best classification rate and the direct method has the poorest. Changes in the amount of within-group spectral variability by changing the range of conditional innovation variances, which is the same for each group, do not affect the performance of the cepstral Fisher’s procedures. However, the classification rates of the procedures that ignore

σ_{jk}^2	n_j	N	Cepstral	Cepstral	Cepstral	Kullback-	
			Multitaper	Direct	Smoothed	Chernoff	Leibler
[0.1, 10]	15	250	91.1 (3.5)	87.3 (3.7)	89.4 (3.6)	80.2 (4.9)	79.2 (4.9)
		500	94.0 (2.7)	91.5 (3.4)	93.2 (2.7)	81.1 (4.8)	80.1 (4.6)
		1000	95.6 (2.8)	94.1 (3.1)	95.4 (2.7)	81.4 (4.8)	80.7 (4.7)
	50	250	93.1 (2.2)	90.0 (2.6)	91.8 (2.5)	83.0 (3.5)	81.9 (3.4)
		500	95.5 (1.8)	93.6 (2.2)	94.9 (1.9)	84.1 (3.8)	82.8 (3.7)
		1000	97.0 (1.7)	95.9 (1.7)	96.8 (1.6)	84.7 (3.6)	83.5 (3.6)
	100	250	93.8 (2.0)	90.6 (2.4)	92.3 (2.2)	84.2 (3.3)	82.8 (3.5)
		500	95.7 (1.7)	93.9 (2.0)	95.2 (1.8)	84.8 (3.2)	83.5 (3.3)
		1000	97.2 (1.5)	96.0 (1.6)	96.9 (1.5)	85.3 (3.2)	84.2 (3.5)
[0.3, 3]	15	250	91.1 (3.1)	86.8 (4.3)	89.3 (3.7)	82.7 (3.9)	82.4 (4.0)
		500	93.9 (3.0)	91.6 (3.6)	93.4 (3.0)	83.7 (3.9)	83.2 (4.0)
		1000	95.5 (2.7)	94.1 (2.9)	95.4 (2.7)	84.4 (4.1)	83.9 (4.1)
	50	250	93.1 (2.1)	89.8 (2.7)	91.7 (2.3)	84.9 (3.2)	84.4 (3.4)
		500	95.5 (1.8)	93.7 (2.1)	94.9 (1.9)	85.8 (3.4)	85.3 (3.5)
		1000	97.0 (1.6)	95.7 (1.8)	96.8 (1.6)	86.6 (3.0)	85.9 (3.5)
	100	250	93.7 (2.0)	90.5 (2.5)	92.3 (2.2)	85.4 (2.9)	84.8 (3.0)
		500	95.7 (1.7)	94.0 (2.0)	95.2 (1.8)	86.5 (2.9)	86.0 (3.1)
		1000	97.3 (1.5)	96.1 (1.7)	97.0 (1.5)	87.4 (2.8)	86.8 (3.1)
[0.9, 1.1]	15	250	91.2 (3.3)	86.9 (4.0)	89.6 (3.6)	85.9 (3.2)	85.7 (3.2)
		500	93.9 (3.0)	91.5 (3.4)	93.4 (3.0)	86.6 (3.2)	86.3 (3.3)
		1000	95.8 (2.6)	94.4 (2.7)	95.5 (2.6)	87.6 (3.1)	87.3 (3.2)
	50	250	93.2 (2.2)	89.9 (2.6)	91.8 (2.3)	86.2 (3.0)	86.1 (3.0)
		500	95.5 (1.8)	93.6 (2.2)	95.0 (1.9)	87.1 (2.8)	86.6 (3.0)
		1000	97.1 (1.6)	95.7 (1.8)	96.8 (1.6)	88.1 (2.8)	87.6 (3.0)
	100	250	93.6 (2.0)	90.4 (2.3)	92.3 (2.2)	86.2 (2.9)	85.8 (3.0)
		500	96.0 (1.7)	94.1 (2.0)	95.4 (1.8)	87.6 (2.6)	87.0 (2.9)
		1000	97.3 (1.4)	96.1 (1.6)	97.1 (1.5)	88.4 (2.6)	88.0 (2.9)

Table 1: Mean (standard deviation) of the percent of correctly classified replicates.

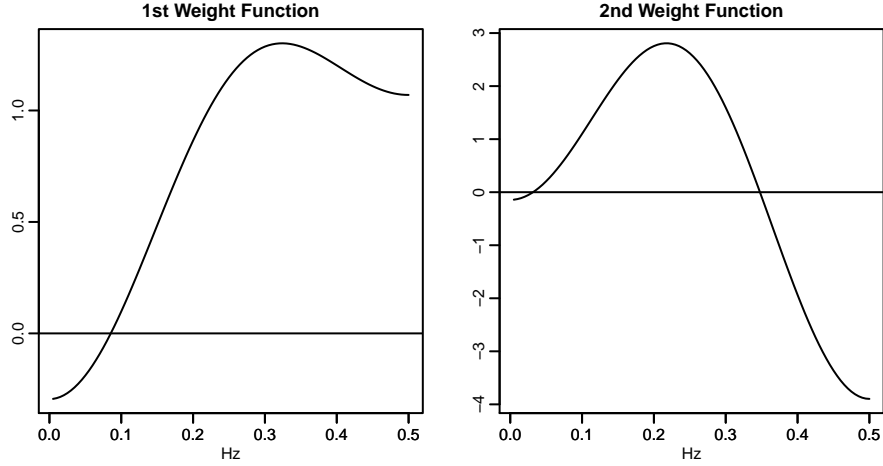


FIGURE 4: Estimated log-spectral weight functions from the gait analysis study.

within-groups spectral variability have reduced classification rates when this variability increases.

6 Analysis of Gait Variability

Patterns of gait variability can provide insight into how neurological conditions affect the systems that regulate walking (Hausdorff et al., 2000; Hausdorff, 2005). The discriminant analysis of gait variability from people with different neurological conditions can provide a tool for characterizing pathologies, aid in the diagnosis of neurodegenerative disease, and aid in the evaluation of treatment efficacy. In this section, we consider a discriminant analysis of gait variability from three groups of participants in the the study described by Hausdorff et al. (2000): healthy controls, participants with amyotrophic lateral sclerosis, a disease characterized by a loss of motoneurons, and participants with Huntington’s disease, a pathology of the basal ganglia. These data can be obtained through *PhysioNet* (Goldberger et al., 2000).

In the study, participants were fitted with pressure sensors on the soles of their feet and told to walk at a normal pace. The information collected was used to compute stride intervals, or the

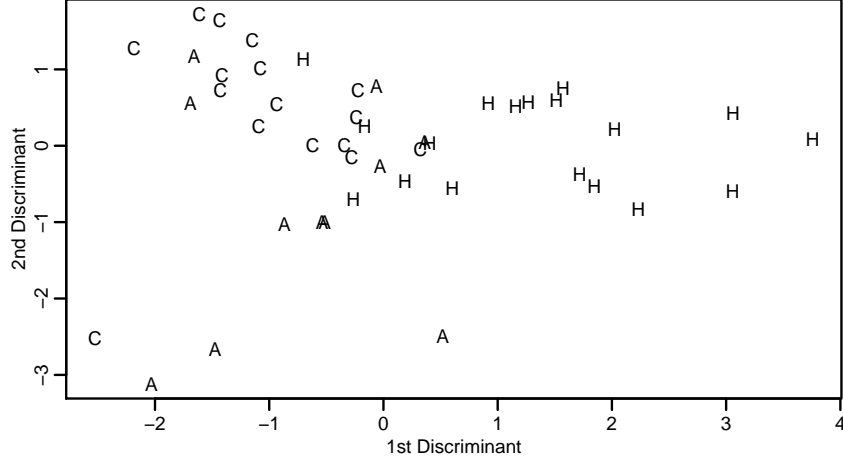


FIGURE 5: Estimated discriminants from the gait analysis study: “C” denotes healthy controls, “A” denotes participants with amyotrophic lateral sclerosis, and “H” denotes participants with Huntington’s disease.

elapsed time for each gait cycle. The present analysis considers the 3.5 minutes of stride intervals defined by the left foot after a 20 second start-up period. After a 3 standard deviation median filter was applied to remove artifacts associated with turning at the end of the hallway (Hausdorff et al., 2000), cubic smoothing spline interpolants of the stride intervals as functions of time were sampled at 2 Hz and linear trends were removed. The resulting data are detrended stride interval series of length $N = 420$ from $n = 45$ participants: 16 healthy controls, 11 participants with amyotrophic lateral sclerosis, and 18 participants with Huntington’s disease.

Log-spectra were estimated using $R = 7$ multitapers and cross-validation selected $L = 4$ coefficients. Estimated log-spectral weight functions $\hat{\xi}_1$ and $\hat{\xi}_2$ are displayed in Figure 4. To aid interpretability, we only display power spectra for frequencies between $[0, 0.5]$ Hz (Henmi et al., 2009).

The first log-spectral weight function is a contrast in power from frequencies greater than 0.08

Hz versus power from frequencies less than 0.08 Hz; large positive values for the first discriminant indicate larger relative power from higher frequencies. The second weight function is a contrast between power from frequencies between 0.05–0.35 Hz versus power from frequencies greater than 0.35 Hz; large positive values for the second discriminant indicate more power from lower frequencies. A scatter plot of the estimated discriminants $\hat{d}_{jk1}, \hat{d}_{jk2}$ is displayed in Figure 5. The first discriminant primarily separates participants with Huntington’s disease from the other two groups while the second discriminant primarily separates controls from participants with amyotrophic lateral sclerosis. This finding that groups are separated by contrasts in power from higher frequencies versus lower frequencies is not unexpected. Neurodegenerative diseases such as amyotrophic lateral sclerosis and Huntington’s disease affect systems that are important in maintaining a steady stride, thus introducing power at higher frequencies. Hausdorff et al. (2000) found that, on average, positive values of the autocorrelation function decay slower for controls as compared to participants with Huntington’s disease, which is indicative of healthy controls possessing more power at lower frequencies, and that the average decay for participants with amyotrophic lateral sclerosis is between that of controls and participants with Huntington’s disease.

Leave-out-one cross-validation was used to empirically assess the effectiveness of the classification rule. Fourteen of the 16 controls were correctly classified, 1 misclassified as having amyotrophic lateral sclerosis, and 1 misclassified as having Huntington’s disease. Fifteen of the 18 participants with Huntington’s disease were correctly classified, 2 incorrectly classified as being controls, and 1 misclassified as having amyotrophic lateral sclerosis. The rule did comparatively worse in classifying participants with amyotrophic lateral sclerosis: 5 were correctly classified, 4 classified as controls, and 2 classified as having Huntington’s disease. The Kullback-Leibler and Chernoff information measure classifiers of Kakizawa et al. (1998) were also implemented and their performances evaluated through leave-out-one cross-validation. The two information measure classifiers had identical perfor-

mance. They correctly classified 11 of the 16 controls, 12 of the 18 participants with Huntington’s disease, and 5 of the 11 participants with amyotrophic lateral sclerosis. In addition to providing a more accurate classification rule, cepstral Fisher’s discriminant analysis produced interpretable estimated log-spectral weight functions and two-dimensional discriminant plot that can be used for illustration, which is lacking from the information measure classifiers.

7 Discussion

Traditional models for the frequency domain discriminant analysis of time series assume each series from a group are independently and identically distributed as models that are used in the spectral analysis of a single time series. However, such models fail to account for within-group spectral variability that is present in most real world applications. This article introduced the stochastic transfer function model that can account for this within-group variability, which is a product of long-range higher-order dependence. A cepstral based Fisher’s discriminant analysis is developed under this model. The procedure provides parsimonious low-dimensional measures that illuminate scientific mechanisms that best separate groups and, when within-group spectral variability is present, provides more accurate classification of new observations as compared to methods that do not account for within-group spectral variability.

The cepstral Fisher’s discriminant analysis was presented in this article under the assumption of equal within-group spectral covariance functions. If these covariance functions differ such that $\Gamma_j(\ell, m) = E(b_{j\ell}b_{jm})$ depends on j , the discriminants can be defined by using $\Gamma = \sum_{j=1}^J \pi_j \Gamma_j$. Under this setting, if group membership is viewed as a random variable such that a randomly selected time series has probability π_j from being in group j , discriminants have the interpretation of being linear functions that maximize the variance of the conditional expected value relative to the expected value of the conditional variance, where conditioning is taken with respect to group

membership. The proposed estimator, which already utilizes the estimate $\hat{\mathbf{\Gamma}}_L = \sum_{j=1}^J \hat{\pi}_j \hat{\mathbf{\Gamma}}_{Lj}$, is still valid in this setting and the consistency results established in Theorem 1 hold. The only statements made in this article that need to be adjusted when log-spectral covariances depends on group are those that are made with respect to classification rates. The optimal classification rule under normality in this settings will be quadratic and, while providing some gain in classification accuracy, fails to address the discriminant analysis question or provide interpretable parsimonious measures, such as the estimated weight functions and discriminants in Figures 4 and 5. Additionally, although the theoretically optimal classifier is quadratic, Fisher’s classification procedures are robust to heteroscedasticity (O’Neill, 1992) and outperform quadratic procedures in practice for high-dimensional data where parameters must be estimated (Cheng, 2004).

Only univariate time series were considered in this article while the information criteria based methods of Kakizawa et al. (1998) are applicable to vector-valued time series. Although the stochastic transfer function model for time series in the presence of within-group spectral variability can be extended to the vector-valued setting in a straight forward manner, due to the non-convexity of the matrix exponential discussed in Section 6 of Krafty and Collinge (2013), multivariate analogues of cepstral coefficients have yet to be developed. Consequently, the extension of the proposed discriminant analysis procedure to the vector-valued setting is not straight forward. A cepstral discriminant analysis could be conducted on the collection of component-specific cepstral coefficients, however such an analysis ignores coherence between different components. The extension of the proposed procedure to the vector-valued setting will be the focus of future work.

Acknowledgements

This work was supported by grant R01GM113243 from the National Institute of General Medical Sciences, U.S.A. I thank two anonymous reviewers whose insightful comments greatly improved the

article.

Appendix: Proofs

Proof of Theorem 1

To prove Theorem 1, we will make use of two lemmas. Lemma 2, whose proof is provided in the subsequent subsection, formalizes the approximation of the true infinite dimensional weight functions from the weight functions obtained from truncated true cepstra $\mathbf{c}_{jk}^L = (c_{jk0}, \dots, c_{jkL-1})^T$, while Lemma 3 follows directly from Assumptions 2-4 and the orthonormality of the cosine series.

Lemma 2. *Under Assumption 2, for a q th cepstral weight function y_q , there exists a series of q th weight functions \mathbf{y}_q^L for the Fisher's discriminant analysis of \mathbf{c}_{jk}^L such that $\|(\mathbf{y}_q^L, 0, \dots) - y_q\|_\Gamma \rightarrow 0$ as $L \rightarrow \infty$*

Lemma 3. *Under Assumptions 2-4, as $N \rightarrow \infty$, $E(\hat{c}_{jk0} - c_{jk0} - \nu_1)^2 = O(N^{-1})$ and $\sup_{\ell=1, \dots, \lfloor N/2 \rfloor} E(\hat{c}_{j\ell k} - c_{j\ell k})^2 = O(N^{-1})$.*

Define $\mathbf{a}_j^L = (a_{j0}, \dots, a_{jL-1})^T$, $\mathbf{b}_{jk}^L = (b_{jk0}, \dots, b_{jkL-1})^T$, $\mathbf{\Gamma}_L = E\left\{\mathbf{b}_{jk}^L (\mathbf{b}_{jk}^L)^T\right\}$, and $\mathbf{\Lambda}_L = \sum_{j=1}^J \pi_j (\mathbf{a}_j^L - \bar{\mathbf{a}}^L) (\mathbf{a}_j^L - \bar{\mathbf{a}}^L)^T$ where $\bar{\mathbf{a}}^L = \sum_{j=1}^J \pi_j \mathbf{a}_j^L$. Further, let $\|\cdot\|$ be the operator norm on $L \times L$ matrices such that, for an $L \times L$ matrix \mathbf{A} , $\|\mathbf{A}\| = \sup_{\mathbf{y}^T \mathbf{y} = 1} (\mathbf{y}^T \mathbf{A}^T \mathbf{A} \mathbf{y})$. Note that showing $\|\hat{\mathbf{\Gamma}}_L^{-1} \hat{\mathbf{\Lambda}}_L - \mathbf{\Gamma}_L^{-1} \mathbf{\Lambda}_L\| \xrightarrow{p} 0$ implies that $\hat{\mathbf{y}}_q^L \xrightarrow{p} \mathbf{y}_q^L$, and since Lemma 2 established that \mathbf{y}_q^L converges to y_q in $\|\cdot\|_\Gamma$, completes the proof of Theorem 1. To ease exposition, we simplify the notation in this proof by suppressing the dependence of parameters and estimates on L . From Cauchy-Schwarz and basic matrix algebra, it follows that

$$\begin{aligned} \left\| \hat{\mathbf{\Gamma}}^{-1} \hat{\mathbf{\Lambda}} - \mathbf{\Gamma}^{-1} \mathbf{\Lambda} \right\| &\leq \left\| \mathbf{\Gamma}^{-1} (\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}) \hat{\mathbf{\Gamma}}^{-1} \hat{\mathbf{\Lambda}} - \mathbf{\Gamma}^{-1} (\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}) \right\| \\ &\leq \sigma^{-1} \left\| \hat{\mathbf{\Gamma}} - \mathbf{\Gamma} \right\| \left\| \hat{\mathbf{\Gamma}}^{-1} \right\| \left\| \hat{\mathbf{\Lambda}} \right\| + \sigma^{-1} \left\| \hat{\mathbf{\Lambda}} - \mathbf{\Lambda} \right\|. \end{aligned} \quad (\text{A.1})$$

To investigate the decay of $\|\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}\| \leq \sum_{j=1}^J \|\hat{\mathbf{\Gamma}}_j - \mathbf{\Gamma}\|$, note that, since the operator norm is dominated by the Frobenius norm, $\mathbb{E}\|\hat{\mathbf{\Gamma}}_j - \mathbf{\Gamma}\|^2 \leq \sum_{\ell,m=0}^{L-1} \mathbb{E}|\hat{\Gamma}_j(\ell, m) - \Gamma(\ell, m)|^2$. From Lemma 3 and the definition of $\hat{\mathbf{\Gamma}}_j$, letting $\eta_\ell = \sup_j \mathbb{E}|b_{jk\ell}|^4$, there exists a $C > 0$ such that $\mathbb{E}\left\{\hat{\Gamma}_j(\ell, m) - \Gamma(\ell, m)\right\}^2 \leq C\sqrt{\eta_\ell\eta_m}(n^{-1} + N^{-1})$, so $\mathbb{E}\|\hat{\mathbf{\Gamma}}_j - \mathbf{\Gamma}\|^2 \leq C\left(\sum_{\ell=0}^{L-1} \sqrt{\eta_\ell}\right)^2(n^{-1} + N^{-1})$. Since $\Pr\left(\sum_{\ell=1}^\infty \ell^2 |b_{jk\ell}|^2 < \infty\right) = 1$ and $\mathbb{E}|b_{jk\ell}|^4 < \infty$, $\eta_\ell = O(\ell^{-4})$ and $\lim_{L \rightarrow \infty} \left(\sum_{\ell=0}^{L-1} \sqrt{\eta_\ell}\right)^2 < \infty$. It then follows that $\|\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}\| = O_p(n^{-1/2}) + O_p(N^{-1/2})$.

It follows from similar arguments to those used above that $\|\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}\| = O_p(n^{-1/2}) + O_p(N^{-1/2})$. Since the approximation of the finite between-group kernel is stable and the infinite dimensional kernel is bounded, $\|\mathbf{\Lambda}\| = O_p(1)$ (Chatelin, 1981, Proposition 2.2), and consequently, $\|\hat{\mathbf{\Lambda}}\| = O_p(1)$.

To investigate the decay of $\|\hat{\mathbf{\Gamma}}^{-1}\|$, let $\hat{\sigma}$ be the smallest eigenvalue of $\hat{\mathbf{\Gamma}}$ so that $\|\hat{\mathbf{\Gamma}}^{-1}\| = \hat{\sigma}^{-1}$. Define the matrix $\tilde{\mathbf{\Gamma}} = \mathbf{\Gamma}^{-1/2}\hat{\mathbf{\Gamma}}\mathbf{\Gamma}^{-1/2}$, let $\tilde{\sigma}$ be its smallest eigenvalue, and note that $\hat{\sigma}^{-1} \leq \sigma^{-1}\tilde{\sigma}^{-1}$. It follows from Lemma 3 and from Bai and Yin (1993) that $\tilde{\sigma} = 1 + O_p(Ln^{-1}) + O_p(N^{-1/2})$, and consequently when $N \rightarrow \infty$ and $Ln^{-1/2} \rightarrow 0$, $\|\hat{\mathbf{\Gamma}}^{-1}\| = \sigma^{-1} + O_p(\sigma^{-1}Ln^{-1}) + O_p(\sigma^{-1}N^{-1/2})$.

Plugging these results for $\|\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|$, $\|\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}\|$, $\|\hat{\mathbf{\Lambda}}\|$ and $\|\hat{\mathbf{\Gamma}}^{-1}\|$ into Equation A.1, we conclude that $\|\hat{\mathbf{\Gamma}}^{-1}\hat{\mathbf{\Lambda}} - \mathbf{\Gamma}^{-1}\mathbf{\Lambda}\| = O_p(\sigma^{-2}n^{-1/2}) + O_p(\sigma^{-2}N^{-1/2}) + O_p(\sigma^{-2}Ln^{-1})$ which, when $\sigma^{-2}n^{-1/2} \rightarrow 0$, $\sigma^{-2}N^{-1/2} \rightarrow 0$ and $Ln^{-1/2} \rightarrow 0$, decays to zero.

Proof of Lemma 2

For an $\mathbb{R}^{\mathbb{N}}$ -valued random variable g with covariance kernel Γ , let $L^2(g)$ represent the Hilbert space spanned by linear functions of g with inner product $\langle \sum_{\ell=0}^\infty w_\ell g_\ell, \sum_{m=0}^\infty v_m g_m \rangle_g = \sum_{\ell,m=0}^\infty w_\ell v_m \Gamma(\ell, m) = \langle w, v \rangle_\Gamma$. Further, let $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ be the inner product of the reproducing kernel Hilbert space $\mathcal{H}(\Gamma)$ that has reproducing kernel Γ . As discussed in Aronszajn (1950), there exists a congruence between $\mathcal{H}(\Gamma)$ and $L^2(g)$ defined through the linear map $\Psi\{\Gamma(\ell, \cdot)\} = g_\ell$. Define the operator $T : \mathcal{H}(\Gamma) \rightarrow \mathcal{H}(\Gamma)$ as $(Th)_\ell = \langle \Lambda(\ell, \cdot), h \rangle_{\mathcal{H}}$, $h \in \mathcal{H}(\Gamma)$. The operator T is self-adjoint, positive and compact so that it

possess a spectral decomposition $T = \sum_{q=1}^Q \tau_q \theta_q \otimes \theta_q$ where τ_q are nonincreasing eigenvalues and θ_q are orthonormal eigenfunctions in $\mathcal{H}(\Gamma)$ (Shin, 2008, Theorem 2.4). We will also consider $\tilde{\Lambda}_L$ as the minimum $\mathcal{H}(\Gamma) \otimes \mathcal{H}(\Gamma)$ norm interpolate to Λ_L on $(0, \dots, L-1) \times (0, \dots, L-1)$ and its associated operator $T_L : \mathcal{H}(\Gamma) \rightarrow \mathcal{H}(\Gamma)$ defined as $(T_L h)_\ell = \langle \tilde{\Lambda}_L(\ell, \cdot), h \rangle_{\mathcal{H}}$. Let τ_q^L be the nonincreasing set of eigenvalues of the operator T_L and θ_q^L be a set of associated eigenfunctions. The following lemma, whose proof mirrors the proof of Lemma 4 in Eubank and Hsing (2008) that deals with approximating the operator of a canonical correlation analysis, relates the spectral decomposition of T_L to the Fisher's discriminant analysis of \mathbf{c}_{jk}^L .

Lemma 4. *For every q th eigenfunction θ_q^L of T_L , there exists a q th eigenfunction \mathbf{y}_q^L of $\mathbf{\Gamma}_L^{-1} \mathbf{\Lambda}_L$ such that $\|\Psi(\theta_q^L) - \sum_{\ell=0}^{L-1} y_{q\ell}^L g_\ell\|_g = 0$.*

A consequence of Lemma 4 and the $L^2(g)$ equivalence between $\Psi(\theta_q)$ and $\sum_{\ell=0}^{\infty} y_{q\ell} g_\ell$ is that the proof of Lemma 2 can be completed by showing the existence of q th eigenfunctions θ_q^L of T_L such that $\theta_q^L \rightarrow \theta_q$ in $\mathcal{H}(\Gamma)$.

By Cauchy-Schwarz, $\|(T - T_L)h\|_{\mathcal{H}} \leq \|\Lambda - \tilde{\Lambda}_L\|_{\mathcal{H} \otimes \mathcal{H}} \|h\|_{\mathcal{H}}$ for all $h \in \mathcal{H}(\Gamma)$ where $\|\cdot\|_{\mathcal{H} \otimes \mathcal{H}}$ is the tensor product norm. The decay of cepstral coefficients under Assumption 2 implies that $\|\Lambda - \tilde{\Lambda}_L\|_{\mathcal{H} \otimes \mathcal{H}} \rightarrow 0$ and consequently, T_L converges to T in operator norm. This convergence, when combined with the property that T_L is stable, implies that the set of unique eigenvalues of T_L converges to the set of unique eigenvalues of T (Chatelin, 1981, Proposition 2.2). Further, if τ_q^L is a series of eigenvalues that converge to τ_q , the multiplicity of τ_q^L is always greater than or equal to the multiplicity of τ_q (Chatelin, 1981, Lemma 2.1). Since there exists a L_0 such that the rank of T_L is equal to Q for all $L \geq L_0$, for large L , the multiplicity of τ_q^L is the same as τ_q . Proposition 2.3 (iii) of Chatelin (1981) can then be applied to conclude that there exists a set of eigenfunctions θ_q^L of T_L such that $\theta_q^L \rightarrow \theta_q$.

References

- Alagón, J. (1989). Spectral discrimination for two groups of time series. *Journal of Time Series Analysis* 10, 203–214.
- Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Hoboken: Wiley.
- Aronszajn, N. (1950). Theory of reproducing kernels. *American Mathematical Society Transactions* 68, 337–404.
- Bai, Z. D. and Y. Q. Yin (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Annals of Statistics* 21, 1275–1294.
- Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* 60, 217–226.
- Bogert, B. P., M. J. R. Healy, and J. W. Tukey (1963). The quefrency analysis of time series for echoes: cepstrum, pseudo autocovariance, cross-cepstrum and saphe cracking. *Proceedings of the Symposium on Time Series Analysis (M. Rosenblatt, Ed), Chapter 15*, 209–243.
- Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*. Philadelphia: SIAM.
- Chatelin, F. (1981). The spectral approximation of linear operators with applications to computation of eigenelements of differential and integral operators. *SIAM Review* 23, 495–522.
- Cheng, Y. (2004). Asymptotic probabilities of misclassification of two discriminant functions in cases of high dimensional data. *Statistics & Probability Letters* 67, 9 – 17.
- Dargahi-Noubary, G. R. and P. J. Laycock (1981). Spectral ratio discriminants and information theory. *Journal of Time Series Analysis* 16, 201–219.

- Delaigle, A. and P. Hall (2012). Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society, Series B* 74, 267–286.
- Diggle, P. J. and I. Al Wasel (1997). Spectral analysis of replicated biomedical time series. *Applied Statistics* 46, 37–71.
- Eubank, R. L. and T. Hsing (2008). Canonical correlation for stochastic processes. *Stochastic Processes and their Applications* 118, 1634–1661.
- Freyermuth, J.-M., H. Ombao, and R. von Sachs (2010). Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series. *Journal of the American Statistical Association* 105, 634–646.
- Goldberger, A. L., L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* 101.
- Hausdorff, J. M. (2005). Gait variability: methods, modeling, and meaning. *Journal of Neuroengineering and Rehabilitation* 2.
- Hausdorff, J. M., A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger (2000). Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *Journal of Applied Physiology* 88, 2045–2053.
- Henmi, O., Y. Shiba, T. Saito, H. Tsuruta, A. Takeuchi, M. Shirataka, S. Obuchi, M. Kojima, and N. Ikeda (2009). Spectral analysis of gait variability of stride interval time series: comparison of young, elderly and Parkinson’s disease patients. *Journal of Physical Therapy Science* 21, 105–111.

- Iannaccone, R. and S. Coles (2001). Semiparametric models and inference for biomedical time series with extra-variation. *Biostatistics* 2, 261–276.
- Janas, D. and R. von Sachs (1995). Consistency for non-linear functionals of the periodogram of tapered data. *Journal of Time Series Analysis* 16, 585–606.
- Johnson, R. A. and D. W. Wichern (2007). *Applied Multivariate Statistical Analysis* (6 ed.). Upper Saddle River: Prentice Hall.
- Kakizawa, Y., R. H. Shumway, and M. Taniguchi (1998). Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association* 93, 328–340.
- Krafty, R. T. and W. O. Collinge (2013). Penalized multivariate Whittle likelihood for power spectrum estimation. *Biometrika* 100, 447–458.
- Krafty, R. T., M. Hall, and W. Guo (2011). Functional mixed effects spectral analysis. *Biometrika* 98, 583–598.
- Ombao, H., J. Raz, R. L. Strawderman, and R. von Sachs (2001). A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika* 88, 1186–1192.
- O’Neill, T. J. (1992). Error rates of non-Bayes classification rules and the robustness of Fisher’s linear discriminant function. *Biometrika* 79, 177–184.
- Parzen, E. (1961). An approach to time series analysis. *Annals of Mathematical Statistics* 32, 951–989.
- Parzen, E. (1962). Extraction and detection problems and reproducing kernel Hilbert spaces. *Journal of SIAM Control Series A* 1, 35–62.

- Percival, D. B. and A. T. Walden (1993). *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge: Cambridge.
- Saavedra, P., C. N. Hernandez, and J. Artiles (2000). Spectral analysis with replicated time series. *Communications in Statistics - Theory and Methods* 29, 2343–2362.
- Saavedra, P., C. N. Hernandez, I. Luengo, J. Artiles, and A. Santana (2008). Estimation of population spectrum for linear processes with random coefficients. *Computational Statistics* 23, 79–98.
- Shin, H. (2008). An extension of Fisher’s discriminant analysis for stochastic processes. *Journal of Multivariate Analysis* 99, 1191–1216.
- Shumway, R. H. (1982). Discriminant analysis for time series. In P. R. Krishnaiah and L. N. Kanal (Eds.), *Classification, Pattern Recognition and Reduction of Dimensionality, Handbook of Statistics*, Volume 2, pp. 1–46. Amsterdam: North Holland.
- Shumway, R. H. and D. S. Stoffer (2011). *Time Series Analysis and Its Applications: With R Examples* (3 ed.). New York: Springer.
- Shumway, R. H. and A. N. Unger (1974). Linear discriminant functions for stationary time series. *Journal of the American Statistical Association* 69, 948–956.
- Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* 70, 1055–1096.
- Zhang, G. Q. and M. Taniguchi (1994). Discriminant analysis for stationary vector time series. *Journal of Time Series Analysis* 15, 117–126.